Chapitre 1 : Lois multidimensionnelles

Dans tout le chapitre, $(\Omega, \mathscr{A}, \mathbb{P})$ désigne un espace probabilisé,

1 Vecteur aléatoire, loi jointe, lois marginales.

Soit $(E_i)_{1\leq i\leq d}$ une famille d'ensembles, chacun muni d'une tribu $\mathscr{E}_i$. Pour tout $i \in \{1, \cdots, d\}$, on se donne une fonction $X_i : (\Omega, \mathscr{A}) \to (E_i, \mathscr{E}_i)$.

Définition 1. La famille $X = (X_i)_{1\leq i\leq d} : \Omega \to \times_{1\leq i\leq d} E_i$ est **un vecteur aléatoire** si quelque soit $i \in \{1, \cdots, d\}$, X_i est une variable aléatoire à valeurs dans E_i.

Définition 2. La **tribu produit** $\mathscr{E} = \otimes_{i\in I}\mathscr{E}_i$ sur $E = \times_{1\leq i\leq d} E_i$ est la plus petite tribu sur E qui rend ces deux points équivalents pour toute famille $(X_i)_{1\leq i\leq d}$:

1. $(X_i)_{1\leq i\leq d}$ est un vecteur aléatoire ;
2. $X = (X_i)_{1\leq i\leq d}$ est une variable à valeurs dans $(E, \mathscr{E})$ au sens de la définition (14) du Chapitre 0.

C'est la plus petite tribu contenant tous les ensembles de la forme $\times_{1\leq i\leq d} A_i$, où $A_i \in \mathscr{E}_i$ pour tout $i \in \{1, \cdots, d\}$.

Définition 3. La **loi jointe** du vecteur aléatoire X est la loi de X vu comme une variable aléatoire à valeurs dans $(E, \mathscr{E})$. Les **lois marginales** de X sont les lois des X_i, pour $i \in I$.

Proposition 4. *Si on connait la loi jointe μ de X, alors on connait toutes les lois marginales μ_i de X. La réciproque est fausse.*

1.1 Cas discret.

Dans cette sous-sous section, les E_i sont finis ou dénombrables. Par conséquent, E est fini ou dénombrable, et la loi de X est caractérisée par sa fonction de masse $f : E \to [0; 1]$ définie pour tout $x = (x_1, \ldots, x_d) \in E$ par

$$f(x_1, \ldots, x_d) = \mathbb{P}(X_1 = x_1, \ldots, X_d = x_d)\,.$$

Proposition 5. Lois marginales dans le cas discret.
Fixons j entre 1 et d. Si f est la fonction de masse X, et f_j la fonction de masse de X_j, alors, pour tout $x_j \in E_j$, on a

$$f_j(x_j) = \sum_{x_1\in E_1} \cdots \sum_{x_{j-1}\in E_{j-1}} \sum_{x_{j+1}\in E_{j+1}} \cdots \sum_{x_d\in E_d} f(x_1, \cdots, x_d).$$

1.2 Cas réel.

Dans cette sous-sous section, on suppose que tous les E_i sont égaux à $\mathbb{R}$ et $E = \mathbb{R}^d$.

Définition 6. La **fonction de répartition** de X est la fonction $F : \mathbb{R}^d \to [0; 1]$ définie pour tout $(x_1, \ldots, x_d) \in \mathbb{R}^d$ par $F(x_1, \ldots, x_d) = \mathbb{P}(X_1 \leq x_1, \ldots, X_d \leq x_d)$.

Proposition 7. Fonction de répartition des lois marginales.
Fixons j entre 1 et d. Soit F la fonction de répartition de X et F_j celle de X_j. Alors, pour tout $x_j \in \mathbb{R}$,

$$F_j(x_j) = \lim_{x_1\to+\infty} \cdots \lim_{x_{j-1}\to+\infty} \lim_{x_{j+1}\to+\infty} \cdots \lim_{x_d\to+\infty} F(x_1, \cdots, x_d)$$

et ces limites commutent.

Définition 8. La loi μ admet pour **densité jointe** la fonction $f : \mathbb{R}^d \to [0; +\infty[$ si, pour tout A borélien de $\mathbb{R}^d$,

$$\mu(A) = \int\!\!\cdots\!\!\int_{x\in A} f(x)\,\mathrm{d}x.$$

Proposition 9. *Si μ admet une densité f, et si F est la fonction de répartition de μ, alors p.p. $f = \dfrac{\partial^d F}{\partial x_1 \cdots \partial x_d}$.*

Proposition 10. Densités marginales.
Fixons j entre 1 et d. Si la loi de X admet pour densité jointe f, alors la loi de X_j admet une densité f_j, dite densité marginale, et pour (presque) tout $x_j \in \mathbb{R}$,

$$f_j(x_j) = \int\!\!\cdots\!\!\int_{(x_1,\ldots,x_{j-1})\in\mathbb{R}^{j-1}} \int\!\!\cdots\!\!\int_{(x_{j+1},\ldots,x_d)\in\mathbb{R}^{d-j}} f(x)\,\mathrm{d}x_1 \cdots \mathrm{d}x_{j-1}\ \mathrm{d}x_{j+1} \cdots \mathrm{d}x_d.$$

1.3 Autres cas

Dans un vecteur aléatoire, certaines coordonnées peuvent être des variables aléatoires discrètes, et d'autres des variables aléatoires réelles,... Exemples. Variables aléatoires complexes...

1.4 Calculs effectifs

Il arrive souvent qu'on s'intéresse à la loi d'un vecteur Y défini à partir d'un vecteur X dont on connait la loi, par une relation $Y = \phi(X)$. On peut alors tenter de calculer la fonction de répartition, ou la fonction caractéristique de Y, ces deux fonctions caractérisant la loi de Y. De façon plus générale, on peut tenter d'exprimer $\mathbb{E}(h(Y))$ pour une fonction h mesurable bornée (méthode de la fonction muette). Si X est de densité f , on a

$$\mathbb{E}(h(Y)) = \mathbb{E}(h \circ \phi(X)) = \int \cdots \int_{\mathbb{R}^d} h \circ \phi(x) f(x) dx_1 \cdots dx_d \,.$$

On cherche alors à exprimer cette intégrale sous la forme $\int h(y)g(y)dy$. Si cela est possible, g est la densité de Y. Dans ce cadre, on rappelle la formule du changement de variables.

Proposition 11. Formule du changement de variables.
Soit $\phi : \mathbb{R}^d \to \mathbb{R}^d$ un difféomorphisme (i.e. une bijection différentiable et d'inverse ϕ^{-1} différentiable) qui à $x = (x_1, \cdots, x_d)$ associe $\phi(x) = (\phi_1(x), \cdots, \phi_d(x))$. On note $J_\phi(x)$ la matrice jacobienne de ϕ définie par

$$J_\phi(x) = \begin{pmatrix} \frac{\partial \phi_1}{\partial x_1}(x) & \cdots & \frac{\partial \phi_d}{\partial x_1}(x) \\ \vdots & \cdots & \vdots \\ \frac{\partial \phi_1}{\partial x_d}(x) & \cdots & \frac{\partial \phi_d}{\partial x_d}(x) \end{pmatrix} .$$

Alors, on a

$$\int \cdots \int_{\mathbb{R}^d} h \circ \phi(x) dx_1 \cdots dx_d = \int \cdots \int_{\mathbb{R}^d} h(y) \frac{1}{|\det(J_\phi)(\phi^{-1}(y))|} dy_1 \cdots dy_d \,.$$

2 Variance, matrice de covariance.

2.1 Variance, covariance.

Soit X et Y deux variables aléatoires réelles telle que $\mathbb{E}(X^2) < +\infty$ et $\mathbb{E}(Y^2) < +\infty$ (on dit alors que X et Y sont de **carré intégrable**).

Définition 12.

$$\mathrm{var}(X) := \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 \,.$$

$$\mathrm{cov}(X, Y) := \mathbb{E}\left[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right] = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \,.$$

Proposition 13. Propriétés de la variance.

1. *$var(X) \geq 0$.*
2. *$var(X) = 0$ ssi X est p.s. constante (égale à $\mathbb{E}(X)$).*
3. *Pour tout $m \in \mathbb{R}$, $var(X) \leq \mathbb{E}\left((X - m)^2\right)$. Ainsi $\mathbb{E}(X)$ est la meilleure prédiction (au sens du risque quadratique) que l'on peut faire de la variable X.*
4. *Pour tout réel α, $var(\alpha X) = \alpha^2 var(X)$.*
5. *$var(X + Y) = var(X) + var(Y) + 2cov(X, Y)$.*

2.2 Matrice de covariance.

Soit $X = (X_1, \cdots, X_d)$ un vecteur aléatoire tel que $\mathbb{E}(\|X\|^2) < +\infty$.

Définition 14. La matrice de covariance de X est la matrice carrée de dimension d, notée Γ, définie par

$$\forall i, j \in \{1, \cdots, d\} \,, \;\; \Gamma_{i,j} = \mathrm{cov}(X_i, X_j) \,.$$

Proposition 15. Propriétés de la matrice de covariance.

1. *Pour tout vecteur* $\alpha \in \mathbb{R}^d$, $\alpha^T \Gamma \alpha = \mathbb{E}\left[\left(\sum_{i=1}^d \alpha_i (X_i - \mathbb{E}(X_i))\right)^2\right]$.
2. Γ *est une matrice symétrique positive.*
3. Γ *admet une valeur propre nulle ssi les variables* X_i *sont linéairement dépendantes, i.e. il existe des réels* $\alpha_0, \alpha_1, \cdots, \alpha_d$ *tels que* $\alpha_0 + \alpha_1 X_1 + \cdots + \alpha_d X_d = 0$ *p.s.*
4. *Si* A *est une matrice* $l \times d$, *et* $Y = AX$, *la matrice de covariance de* Y *est la matrice carrée de dimension* l *donnée par* $A\Gamma\, {}^t\!A$.

3 Indépendance.

3.1 Indépendance d'évènements.

Définition 16. Deux évènements A et B sont **indépendants** ssi $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ (ou de façon équivalente lorsque $\mathbb{P}(B) > 0$, $\mathbb{P}(A|B) = \mathbb{P}(A)$).

Définition 17. Une famille d'évènements $(A_i)_{i \in I}$ est une famille d'évènements indépendants ssi pour toute sous-famille d'indices $J \subset I$ de cardinal fini, on a $\mathbb{P}(\cap_{j \in J} A_j) = \prod_{j \in J} \mathbb{P}(A_j)$.

3.2 Indépendance de variables aléatoires.

On se donne deux variables aléatoires X et Y définies sur $(\Omega, \mathscr{A}, \mathbb{P})$ et à valeurs respectives dans $(E, \mathscr{E})$ et $(F, \mathscr{F})$.

Définition 18. X et Y sont **indépendantes** ssi pour tout $A \in \mathscr{E}$ et tout $B \in \mathscr{F}$, $\mathbb{P}(X \in A; Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$.

Proposition 19. Les fonctions (mesurables) de variables indépendantes sont des variables indépendantes.
Soit f *une fonction mesurable*[1] *de* $(E, \mathscr{E})$ *à valeurs dans* $(E', \mathscr{E}')$. *Soit* g *une fonction mesurable de* $(F, \mathscr{F})$ *à valeurs dans* $(F', \mathscr{F}')$. *Si* X *et* Y *sont indépendantes, alors* $f(X)$ *et* $g(Y)$ *sont indépendantes.*

Proposition 20. Cas des variables discrètes. Lecture de l'indépendance sur la fonction de masse.
Si les ensembles E *et* F *sont de cardinal fini ou dénombrables, les variables* X *et* Y *sont indépendantes ssi pour tout* $x \in E$ *et tout* $y \in F$, $\mathbb{P}(X = x; Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$ *(la fonction masse du couple est le produit des fonctions masse de chacune des variables).*

Proposition 21. Cas des variables réelles. Indépendance et non corrélation.
On suppose que $E = F = \mathbb{R}$, *et* $\mathscr{E} = \mathscr{F} = \mathscr{B}(\mathbb{R})$, *et que les variables* X *et* Y *sont de carré intégrable. Si* X *et* Y *sont indépendantes,* $cov(X, Y) = 0$. **La réciproque est fausse.**

Proposition 22. Cas des variables réelles. Lecture de l'indépendance sur la fonction de répartition.
On suppose que $E = F = \mathbb{R}$, $\mathscr{E} = \mathscr{F} = \mathscr{B}(\mathbb{R})$. *Les variables* X *et* Y *sont indépendantes ssi pour tout* $x \in \mathbb{R}$ *et tout* $y \in \mathbb{R}$, $\mathbb{P}(X \leq x; Y \leq y) = \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y)$ *(la fonction de répartition du couple est le produit des fonctions de répartition de chacune des variables).*

Proposition 23. Cas des variables réelles. Lecture de l'indépendance sur la fonction caractéristique.
On suppose que $E = F = \mathbb{R}$, *et* $\mathscr{E} = \mathscr{F} = \mathscr{B}(\mathbb{R})$. *On note* $\phi_{(X,Y)}$, ϕ_X, ϕ_Y *les fonctions caractéristiques du couple* (X, Y), *et des variables* X *et* Y. *Les variables* X *et* Y *sont indépendantes ssi pour tout* $\theta = (\theta_x, \theta_y) \in \mathbb{R}^2$, $\phi_{(X,Y)}(\theta) = \phi_X(\theta_x)\phi_Y(\theta_y)$ *(la fonction caractéristique du couple est le produit des fonctions caractéristiques de chacune des variables).*

Proposition 24. Cas des variables réelles à densité. Lecture de l'indépendance sur la densité.
On suppose que $E = F = \mathbb{R}$, *et* $\mathscr{E} = \mathscr{F} = \mathscr{B}(\mathbb{R})$, *et que le couple* (X, Y) *est de densité* $h(x, y)$. *Soit* f *et* g *les densités de* X *et* Y *(qui existent d'après la proposition 10). Les variables* X *et* Y *sont indépendantes ssi pour presque tout* $(x, y) \in \mathbb{R}^2$, $h(x, y) = f(x)g(y)$ *(la densité du couple est le produit des densités de chacune des variables).*

Question : Cas où X est discrète et Y est réelle ? ? ?

Tout ce qui précède se généralise à plus de deux variables aléatoires. Donner les énoncés correspondants.

1. Une fonction f de $(E, \mathscr{E})$ dans $(E', \mathscr{E}')$ est dite mesurable ssi pour tout évènement A de E' (i.e. $A \in \mathscr{E}'$), l'ensemble $f^{-1}(A)$ est un évènement de E (i.e. $f^{-1}(A) \in \mathscr{E}$)

4 Vecteurs gaussiens

Définition 25. Soit $X = (X_1, \cdots, X_d)$ un vecteur aléatoire de dimension d. X est un **vecteur gaussien** ssi toute combinaison linéaire des coordonnées de X est une variable réelle gaussienne : pour tout $\alpha = (\alpha_1, \cdots, \alpha_d) \in \mathbb{R}^d$, $\langle \alpha, X \rangle := \alpha_1 X_1 + \cdots + \alpha_d X_d$ est une variable gaussienne.

Proposition 26. Fonction caractéristique d'un vecteur gaussien.
X est un vecteur gaussien ssi il existe $m \in \mathbb{R}^d$ et une matrice symétrique positive Γ de dimension $d \times d$, tels que la fonction caractéristique ϕ_X de X vaut

$$\phi_X(t) = \exp\left(i \langle t, m \rangle\right) \exp\left(-\frac{1}{2} \langle t, \Gamma t \rangle\right).$$

m est alors le vecteur des moyennes de X et Γ sa matrice de covariance :

$$m := \begin{pmatrix} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_d) \end{pmatrix} \; ; \; \Gamma := (cov(X_i, X_j))_{1 \leq i,j \leq d} = \begin{pmatrix} var(X_1) & \cdots & cov(X_1, X_d) \\ \vdots & \ddots & \vdots \\ cov(X_d, X_1) & \cdots & var(X_d) \end{pmatrix}$$

On note $X \sim \mathcal{N}_d(m, \Gamma)$.

Proposition 27. Transformation linéaire d'un vecteur gaussien.
Soit X un vecteur gaussien de dimension d : $X \sim \mathcal{N}_d(m, \Gamma)$. Soit A une matrice de dimension $l \times d$ et b un vecteur de dimension l. Le vecteur $AX + b$ est un vecteur gaussien de loi $\mathcal{N}_l(Am + b, A\Gamma A^T)$.

Proposition 28. Indépendance dans un vecteur gaussien.
Soit $X = (X_1, \cdots, X_d)$ un vecteur gaussien. Pour tous sous-ensembles d'indices I, J de $\{1, \cdots, d\}$, $(X_i)_{i \in I}$ et $(X_j)_{j \in J}$ sont indépendants ssi $cov(X_i, X_j) = 0$ pour tout $i \in I$, et tout $j \in J$.

Proposition 29. Densité d'un vecteur gaussien.
Soit X un vecteur gaussien de dimension d : $X \sim \mathcal{N}_d(m, \Gamma)$. X admet une densité sur $\mathbb{R}^d$ ssi Γ est définie positive. Dans ce cas, la densité de X est donnée par

$$f_X(x) = \frac{1}{\sqrt{2\pi}^n} \frac{1}{\sqrt{\det(\Gamma)}} \exp\left(-\frac{1}{2}(x, \Gamma^{-1} x)\right).$$

5 Conditionnement.

5.1 Loi conditionnelle par rapport à une variable.

Soit (X, Y) un couple de variables aléatoires à valeurs respectives dans $(E, \mathscr{E})$ et

$(F, \mathscr{F})$. Le but de ce chapitre est de donner une réponse à la question : si j'observe la variable Y, que puis-je dire de la distribution de X ?

5.1.1 Cas où Y est une variable discrète.

Pour tout valeur possible y de la variable Y telle que $\mathbb{P}(Y = y) > 0$, la loi conditionnelle de X sachant que $Y = y$ est la probabilité sur $(E, \mathscr{E})$ qui à un évènement $A \in \mathscr{E}$ associe

$$\mathbb{P}(X \in A | Y = y) = \frac{\mathbb{P}(X \in A ; Y = y)}{\mathbb{P}(Y = y)}. \tag{1}$$

Elle dépend de la valeur fixée y.

Si X est une variable discrète, elle est caractérisée par sa distribution conditionnelle qui à $x \in E$ associe $\mathbb{P}(X = x | Y = y)$. On a

$$\mathbb{P}(X = x | Y = y) = \frac{\mathbb{P}(X = x ; Y = y)}{\sum_{z \in E} \mathbb{P}(X = z ; Y = y)}. \tag{2}$$

5.1.2 Cas où (X, Y) admet une densité.

Dans ce cas, la variable Y admet aussi une densité et pour tout $y \in \mathbb{R}$, $\mathbb{P}(Y = y) = 0$. De même pour tout évènement $A \in \mathscr{B}(\mathbb{R})$ et tout $y \in \mathbb{R}$, $\mathbb{P}(X \in A; Y = y) = 0$. Ainsi, l'expression (1) est une forme indéterminée du type "0/0". Pour définir la probabilité conditionnelle, on doit utiliser un passage à la limite, et définir "$\mathbb{P}(X \in A|Y = y)$" comme la limite quand ϵ tend vers 0 de $\mathbb{P}(X \in A|Y \in [y - \epsilon; y + \epsilon])$. Si $f(x, y)$ désigne la densité de (X, Y), on a alors

$$\mathbb{P}(X \in A|Y = y) = \frac{\int_{x \in A} f(x, y)\, dx}{\int_{\mathbb{R}} f(x, y) dx}.$$

Ainsi, la densité conditionnelle de X sachant $Y = y$ est donnée par l'expression (à rapprocher de (2))

$$f_{X|Y=y}(x) = \frac{f(x, y)}{\int_{\mathbb{R}} f(x, y) dx}.$$

5.1.3 Indépendance et conditionnement.

Proposition 30. *Les variables X et Y sont indépendantes ssi pour tout $y \in F$, la loi conditionnelle de X sachant $Y = y$ ne dépend pas de y :*

$$\forall A \in \mathscr{E}, \mathbb{P}(X \in A|Y = y) = \mathbb{P}(X \in A);$$

5.2 Espérance conditionnelle

Soit X une v.a. ou un vecteur aléatoire et Y une v.a. réelle. On cherche à définir l'espérance conditionnelle de Y sachant X comme une moyenne de Y à X fixé. Voici la définition générale.

Définition 31. On suppose que $\mathbb{E}(|Y|) < +\infty$. L'espérance conditionnelle de Y sachant X, notée $\mathbb{E}(Y|X)$, est une variable aléatoire Z intégrable telle que

1. il existe une fonction (mesurable) η telle que $Z = \eta(X)$,
2. pour toute fonction ζ mesurable bornée,
$$\mathbb{E}\left[Y\zeta(X)\right] = \mathbb{E}\left[Z\zeta(X)\right].. \tag{3}$$

Proposition 32. *Deux variables aléatoires Z_1 et Z_2 qui vérifient les deux conditions ci-dessus sont égales presque sûrement. La fonction η est unique (à un presque partout près) sur le support de la loi de X. On note alors $\eta(x)$ par*

$$\mathbb{E}(Y|X = x).$$

Attention, bien souvent ici, $\mathbb{P}(X = x) = 0$ et donc $\mathbb{E}(Y|X = x) \neq \dfrac{\mathbb{E}\Big(Y\mathbf{1}\{X = x\}\Big)}{\mathbb{P}(X = x)}$. On peut vérifier en revanche que cette formule est vraie si la loi de X est discrète (voir proposition ci-dessous).

Si A est un événement, $\mathbb{P}(A|X)$ est une espérance conditionnelle de $Y = \mathbf{1}_A$ sachant X, i.e..

$$\mathbb{P}(A|X) = \mathbb{E}(\mathbf{1}_A|X) \quad \text{p.s.}$$

La définition ci-dessus n'est pas intuitive. Mais si on suppose que Y est L^2, on obtient une caractérisation de $\mathbb{E}(Y|X)$ plus claire.

Proposition 33. Caractérisation dans le cas L^2.
Si $\mathbb{E}(Y^2) < \infty$, alors $\mathbb{E}(Y|X)$ est la meilleure approximation de Y par une fonction de X au sens des moindres carrés. Autrement dit, quelque soit la variable aléatoire $Z' = \eta'(X)$ de carré intégrable, $\mathbb{E}\left[(Y - E(Y|X))^2\right] \leq \mathbb{E}\left[(Y - \eta'(X))^2\right]$

Proposition 34. Calcul dans le cas où X est discrète.
On suppose que l'ensemble $\{x_i, i \in I\}$ des valeurs prises par X est de cardinal fini ou dénombrable. Alors,

$$\mathbb{E}(Y|X) = \sum_{i \in I} \mathbb{E}(Y|X = x_i)\ \mathbf{1}_{X=x_i}, \quad \text{où } \mathbb{E}(Y|X = x_i) = \frac{\mathbb{E}(Y\ \mathbf{1}_{X=x_i})}{\mathbb{P}(X = x_i)}.$$

Proposition 35. Calcul dans le cas où (X, Y) a une densité.
On suppose que le couple (X, Y) est de densité $f(x, y)$ sur $\mathbb{R}^2$. On note g la densité de la loi marginale de X, autrement dit $g(x) = \int f(x, y)\,\mathrm{d}y$ pour tout x. On a alors

$$\mathbb{E}(Y|X) = \int_{\mathbb{R}} y \frac{f(X, y)}{g(X)}\,\mathrm{d}y \quad p.s.\,.$$

Autrement dit, quelque soit x,

$$\mathbb{E}(Y|X = x) = \int_{\mathbb{R}} y \frac{f(x, y)}{g(x)}\,\mathrm{d}y = \int_{\mathbb{R}} y f_{Y|X=x}(y)\,\mathrm{d}y.$$

Théorème 36. Désintégration.
Soit Y est une v.a. indépendante de X, et soit T est une variable aléatoire fonction de X. Soit φ une fonction mesurable. Si $\varphi(T, Y)$ est intégrable, alors

$$\mathbb{E}\Big(\varphi(T, Y)\Big|X\Big) = \Phi(T) \quad p.s. \quad \text{où, pour tout } t, \quad \Phi(t) = \mathbb{E}\big(\varphi(t, Y)\big).$$

Proposition 37. Propriétés de l'espérance conditionnelle de Y sachant X.
Dès que les variables aléatoires à l'intérieur des espérances conditionnelles sont intégrables, on a

1. $\mathbb{E}\Big(\mathbb{E}(Y|X)\Big) = \mathbb{E}(Y)$
2. *si $Y \geq 0$ p.s., alors $\mathbb{E}(Y|X) \geq 0$ p.s. ; si $Y_1 \leq Y_2$ p.s., alors $\mathbb{E}(Y_1|X) \leq \mathbb{E}(Y_2|X)$*
3. *si α et β sont deux constantes, $\mathbb{E}(\alpha Y_1 + \beta Y_2|X) = \alpha\mathbb{E}(Y_1|X) + \beta\mathbb{E}(Y_2|X)$ p.s.*
4. *si X et Y sont indépendantes, $\mathbb{E}(Y|X) = \mathbb{E}(Y)$ p.s.*
5. *(Jensen conditionnel) si φ est une fonction convexe telle que $\varphi(Y)$ soit intégrable, alors*

$$\varphi\Big(\mathbb{E}(Y|X)\Big) \leq \mathbb{E}\Big(\varphi(Y)\Big|X\Big) \quad p.s.$$

6. *si T est une fonction de X, alors $\mathbb{E}(TY|X) = T\,\mathbb{E}(Y|X)$ p.s.*
7. *(Beppo-Levi conditionnel) si (Y_n) est une suite croissante de v.a. positives, et si $Y_\infty = \lim_n Y_n$, alors*

$$\lim_n \mathbb{E}(Y_n|X) = \mathbb{E}(Y_\infty|X) \quad p.s.$$

8. *(Fatou conditionnel) si (Y_n) est une suite de v.a. positives, alors $\mathbb{E}\Big(\liminf_n Y_n\Big|X\Big) \leq \liminf_n \mathbb{E}(Y_n|X)$ p.s.*
9. *(Lebesgue conditionnel) si (Y_n) est une suite de v.a. qui converge p.s. vers Y_∞ et s'il existe Z intégrable telle que pour tout n, $|Y_n| \leq Z$, alors,*

$$\mathbb{E}\Big(\big|Y_n - Y_\infty\big|\Big|X\Big) \to 0 \quad p.s.$$

5.3 Conditionnement dans un vecteur gaussien.

Proposition 38. *Soit X et Y deux vecteurs aléatoires (de dimension respective n et d) tels que (X, Y) est un vecteur gaussien de dimension $n+d$ de vecteur des moyennes $\begin{pmatrix} m_X \\ m_Y \end{pmatrix}$ et de matrice de covariance $\Gamma = \begin{pmatrix} \Gamma_{XX} & \Gamma_{XY} \\ \Gamma_{YX} & \Gamma_{YY} \end{pmatrix}$. On suppose que Γ_{XX} est définie positive. La loi conditionnelle du vecteur Y sachant X est la loi d'un vecteur gaussien de moyenne $m_Y + \Gamma_{YX}\left(\Gamma_{XX}\right)^{-1}(X - m_x)$ et de matrice de covariance $\Gamma_{YY} - \Gamma_{YX}\left(\Gamma_{XX}\right)^{-1}\Gamma_{XY}$.*

En particulier, $\mathbb{E}(Y|X) = m_Y + \Gamma_{YX}\left(\Gamma_{XX}\right)^{-1}(X - m_x)$ est une transformation affine de X.

6 Convergence des moyennes empiriques.

Les deux théorèmes de cette section donnent des informations sur le comportement quand n tend vers l'infini de moyennes empiriques. Aussi, ils sont au coeur de la statistique dans la limite des grands échantillons. Dans toute la suite, on se donne une suite de vecteurs aléatoires $(X_n)_{n\geq 1}$ indépendants et de même loi ("i.i.d." pour "indépendants identiquement distribués") à valeurs dans $\mathbb{R}^d$. On suppose que $\mathbb{E}(\|X_1\|) < +\infty$ et on note m le vecteur des moyennes : $m = \mathbb{E}(X_1)$. Lorsque $\mathbb{E}(\|X_1\|^2) < +\infty$, on peut aussi définir la matrice Γ de covariance de X_1. On pose

$$S_n := X_1 + \cdots + X_n \in \mathbb{R}^d , \quad \text{et} \quad \bar{X}_n := \frac{1}{n}(X_1 + \cdots + X_n) = \frac{S_n}{n} \in \mathbb{R}^d$$

la moyenne empirique des X_i.

Proposition 39. Moyenne et matrice de covariance de la moyenne empirique.

1. *Si* $\mathbb{E}(\|X_1\|) < +\infty$, *on a* $\mathbb{E}(\bar{X}_n) = m$.
2. *Si* $\mathbb{E}(\|X_1\|^2) < +\infty$, *la matrice de covariance* $\bar{\Gamma}_n$ *de* $\bar{X}_n$ *est égale à* Γ/n : $\boxed{\bar{\Gamma}_n = \frac{1}{n}\Gamma}$.
   En particulier,
   $$\mathbb{E}\left[\left\|\bar{X}_n - m\right\|^2\right] = \mathit{Trace}(\bar{\Gamma}_n) = \frac{1}{n}\mathit{Trace}(\Gamma) ,$$
   et
   $$\bar{X}_n \xrightarrow[n\to+\infty]{L^2} m .$$

Ainsi, la moyenne empirique converge en moyenne quadratique vers la moyenne théorique. Ce dernier résultat peut être considérablement amélioré :

Théorème 40. Loi forte des grands nombres.
On suppose que $\mathbb{E}(\|X_1\|) < +\infty$. *Alors* $\boxed{\bar{X}_n \xrightarrow[n\to+\infty]{p.s.} m}$.

Ainsi, en dehors d'un ensemble de probabilité nulle, une réalisation de la moyenne empirique (i.e. le calcul de la moyenne empirique sur un tirage de notre échantillon) converge vers la moyenne théorique.

Le résultat suivant précise à quelle vitesse a lieu la convergence de la moyenne empirique vers la moyenne théorique, et quelle est la distribution statistique des fluctuations de la moyenne empirique autour de la moyenne théorique. Il permet par conséquent de construire des intervalles de confiance pour la moyenne théorique.

Théorème 41. Théorème central limite.
On suppose que $\mathbb{E}(\|X_1\|^2) < +\infty$. *Alors,*

$$\boxed{\sqrt{n}\left(\bar{X}_n - m\right) \xrightarrow[n\to+\infty]{loi} Z \sim \mathscr{N}(0, \Gamma) .}$$

Le théorème central limite se généralise à des fonctions suffisamment régulières de la moyenne empirique :

Corollaire 42. Soit f une fonction de $\mathbb{R}^d$ à valeurs de $\mathbb{R}$, de classe $\mathscr{C}_b^1$. Alors

$$\sqrt{n}\left(f(\bar{X}_n) - f(m)\right) \xrightarrow[n\to+\infty]{loi} Z \sim \mathscr{N}(0, \langle \nabla f(m), \Gamma \nabla f(m)\rangle) .$$

7 Exercices

Exercice 1. Quelle est la densité du couple (X, Y) dont la fonction de répartition jointe est donnée par

$$F(x,y) = \begin{cases} 0 & \text{si } x < 0, \\ (1 - e^{-x})\left(\frac{1}{2} + \frac{1}{\pi}\arctan(y)\right) & \text{si } x \geq 0. \end{cases}$$

Exercice 2. Pour les deux fonctions F définies ci-dessous, laquelle ou lesquelles sont des fonctions de répartition de couples (X, Y) ?

$$F(x,y) = \begin{cases} 1 - e^{-x-y} & \text{si } x, y \geq 0, \\ 0 & \text{sinon.} \end{cases} \qquad F(x,y) = \begin{cases} 1 - e^{-x} - xe^{-y} & \text{si } 0 \leq x \leq y, \\ 1 - e^{-y} - ye^{-x} & \text{si } 0 \leq y \leq x, \\ 0 & \text{sinon.} \end{cases}$$

Si la réponse est oui, quelles sont les deux fonctions de répartitions marginales ?

Exercice 3. Soit (X, Y, Z) le triplet aléatoire réel dont la densité est définie, lorsqu'elle est non nulle, par

$$f(x, y, z) = (y - x)^2 \exp\Big(- (1 + z)(y - x)\Big) \text{ si } 0 \leq x \leq 1,\ y \geq x \text{ et } z \geq 0.$$

On note $U = X$, $V = (Y - X)$ et $W = Z(Y - X)$. Quelle est la loi de (U, V, W) ?

Exercice 4. Soit $a > 0$ et $0 < p < 1$. Soit X une variable aléatoire de loi de Poisson $\mathscr{P}(a)$. Soit Y une variable aléatoire à valeurs entières telle que, pour tous entiers $0 \leq k \leq n$, $\mathbb{P}(Y = k|X = n) = \binom{n}{k} p^k (1 - p)^{n-k}$. On pose $Z = X - Y$. Quelle est la loi du couple (Y, Z) ?

Exercice 5. On lance deux dés, un rouge et un bleu, et on note les numéros obtenus. Soit les évènements A :"le dé rouge amène un numéro pair", B : " le dé bleu amène un numéro pair", C :"la somme des numéros est paire". Calculer les probabilités de A, B, C. Vérifier que A, B, C sont deux à deux indépendants, mais ne sont pas mutuellement indépendants.

Exercice 6. Soient A et B deux évènements incompatibles. Montrer que A et B sont indépendants si et seulement si $\mathbb{P}(A)$ ou $\mathbb{P}(B)$ est nulle.

Exercice 7. On effectue des essais indépendants de probabilité de succès constante égale à p, $0 < p < 1$ jusqu'à obtenir un nombre m fixé à l'avance de succès. Soit X le nombre d'essais nécessaires.

1. Calculer la loi de probabilité de X
2. Montrer que $\mathbb{E}\left(\frac{m-1}{X-1}\right) = p$ et que $\mathbb{E}\left(\frac{m}{X}\right) \neq p$ (supposer ici $m > 1$)

Exercice 8. Soient X_1, X_2, X_3, X_4 des variables aléatoires indépendantes de loi de Bernoulli $\mathscr{B}(p)$. On considère la matrice $M = \begin{pmatrix} X_1 & X_2 \\ X_3 & X_4 \end{pmatrix}$, et on note D son déterminant. Calculer $\mathbb{E}(D)$.

Exercice 9. On tire au hasard deux numéros de l'ensemble $\{-2, -1, 0, 1, 2\}$, et on note X leur produit. Calculer $\mathbb{E}(X)$ *(1)* lorsque le tirage a lieu avec remise ; et *(2)* lorsque le tirage a lieu sans remise.

Exercice 10. Soit X_1, X_2 deux variables de Bernoulli indépendantes, de paramètres respectifs p_1 et p_2. Soit $Y_i; i = 1, 2$ les variables définies par $Y_i = 2X_i - 1$. Est-ce-que Y_1 et Y_2 sont indépendantes ? Est-ce-que Y_1 et Y_1Y_2 sont indépendantes ?

Exercice 11. Soit X une variable aléatoire de loi uniforme sur l'intervalle $[-2, 1]$. On pose $Y = |X|$, $Z = \max(X, 0)$. Trouvez les fonctions de répartition de Y et Z. Y et Z sont-elles des variables à densité ? Y et Z sont-elles indépendantes ?

Exercice 12. Soit (X, Y) un couple de variables aléatoires à valeurs dans $\{-2, 0, 1\} \times \{-1/2, 0, 1\}$. Le tableau suivant donne la valeur de $\mathbb{P}[X = x; Y = y]$ pour les différentes valeurs de x et y.

$y\backslash x$	- 2	0	1
-1/2	1/10	a	0
0	3/10	0	3/10
1	1/10	1/10	0

1. Quelle est la valeur de a ?
2. Quelle est la loi de Y ?
3. Les variables X et Y sont-elles indépendantes ? Justifiez votre réponse.

Exercice 13. Soit (X, Y) un couple de variables aléatoires de densité $f_{(X,Y)}(x, y) = \exp(-y)\ \mathbf{1}_{0 \leq x \leq y}$.

1. Vérifier que $f_{(X,Y)}$ est bien une densité.
2. Les variables X et Y sont-elles indépendantes ?
3. Les variables X et $Y - X$ sont-elles indépendantes ?
4. Quelle est la loi du couple $(Y - X, X/Y)$?

Exercice 14.

1. Si $X \sim \mathscr{B}(n_1, p)$ et $Y \sim \mathscr{B}(n_2, p)$ sont deux variables indépendantes, quelle est la loi de $X + Y$?
2. Si $X \sim \mathscr{P}(\lambda)$ et $Y \sim \mathscr{P}(\mu)$ sont deux variables indépendantes, quelle est la loi de $X + Y$?

Exercice 15. Deux centraux téléphoniques, indépendants entre-eux, reçoivent par jour un nombre d'appels X et Y, qui suivent des lois de Poisson de paramètres respectifs λ et μ.

1. Quelle est la probabilité que l'ensemble des deux centraux reçoit au plus 3 appels, lorsque $\lambda = 2$ et $\mu = 4$?
2. Quelle est la probabilité que $X = k$ sachant que $X + Y = n$, pour deux entiers k et n ? De quelle loi s'agit-il ?

3. En supposant que $\lambda = 2$ et $\mu = 4$, et en sachant que l'ensemble des deux centraux a reçu 8 appels, quelle est la probabilité que le premier central en ait reçu k ? Pour quelle valeur de k cette probabilité conditionnelle est maximale ?

Exercice 16. Soit (X, Y) un couple de variables aléatoires dont la loi est définie par le tableau

X \ Y	0	1	2
0	1/9	2/9	0
1	0	1/9	2/9
2	2/9	0	1/9

1. Montrer que $\text{cov}(X, Y) = 0$, mais que X et Y ne sont pas indépendantes.
2. Déterminer les fonctions génératrices G_X, G_Y, G_{X+Y} de X, Y, $X+Y$, et vérifier que l'on a $G_X G_Y = G_{X+Y}$.
3. Calculer $\mathbb{E}(X|Y)$.

Exercice 17. Soit X une v.a. réelle de loi symétrique (X et $-X$ ont même loi). Soit ε une v.a. indépendante de X telle que $\mathbb{P}(\varepsilon = 1) = p = 1 - \mathbb{P}(\varepsilon = -1)$, pour p dans $]0, 1[$.

1. Donner la loi de εX.
2. A quelle condition sur p, la covariance entre X et εX est-elle nulle ? Dans ce cas, ces deux variables sont-elles indépendantes ?
3. Soit $Y = \mathbf{1}_{X>0} - \mathbf{1}_{X<0}$. Donner la loi de Y, de XY. Calculer la covariance entre $|X|$ et Y. Ces deux v.a. sont-elles indépendantes ?

Exercice 18. Soit X un vecteur aléatoire de dimension 3 de loi $\mathscr{N}(0, \Gamma)$ où

$$\Gamma = \begin{pmatrix} 3 & -1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix} .$$

Trouver un vecteur de la forme $a(X)$ (où a est une application linéaire de $\mathbb{R}^3$ dans $\mathbb{R}^3$), dont les composantes sont indépendantes.

Exercice 19. Soit X une v.a. de loi $\mathscr{N}(0, 1)$, et ϵ une v.a. indépendante de X de loi définie par : $\mathbb{P}(\epsilon = 1) = \mathbb{P}(\epsilon = -1) = 1/2$. Calculer la loi de $Y = \epsilon X$. Le couple (X, Y) est-il gaussien ? Calculer $\text{cov}(X, Y)$.

Exercice 20. 1. Montrer que la matrice $\Gamma = \begin{pmatrix} 1 & \gamma \\ \gamma & 1 \end{pmatrix}$ est une matrice de covariance, si et seulement si $\gamma \in [-1; 1]$.

Dans la suite de l'exercice, on suppose que cette condition est vérifiée, et on considère (X, Y) un vecteur gaussien centré de matrice de covariance Γ.

2. Trouver la valeur a_0 de a qui minimise $\mathbb{E}\left((Y - aX)^2\right)$.
3. Quelle est la loi de X, de $Y - a_0 X$, de $(X, Y - a_0 X)$? Montrer que X et $Y - a_0 X$ sont indépendantes.
4. Calculer $\mathbb{E}(Y|X)$.

Exercice 21. Soit $\Gamma = (\gamma_{ij})$ une matrice carrée de dimension 3 définie par $\gamma_{ii} = 1$, et $\gamma_{ij} = a$ si i et j sont distincts.

1. Pour quelles valeurs de a, Γ est-elle une matrice de covariance ?

On suppose dans la suite que Γ est une matrice de covariance, et soit (X, Y, Z) un vecteur gaussien centré de covariance Γ.

2. Calculer $\mathbb{E}(Z \mid (X, Y))$, puis $\mathbb{E}(Z \mid X + Y)$.
3. Calculer $\mathbb{E}(X^2 Y^2)$ et $\mathbb{E}(X^2 Y^4)$.

Exercice 22. Soient M et X deux v.a.r. On suppose que M est de loi gaussienne et que, pour tout t réel $\mathbb{E}\left(e^{itX} \mid M\right) = \exp\left(itM - (\sigma^2 t^2/2)\right)$. Montrer que (X, M) est un vecteur gaussien et calculer $\mathbb{E}\left(X \mid M\right)$.

Exercice 23. Loi conditionnelle. Cas discret.
Le nombre Y de pannes d'une machine au cours d'une année est lié à l'âge X (en années) de cette machine. Pour un âge x fixé, la distribution de Y est une loi de Poisson de paramètre $\mu_Y(x) = 1 + \ln(x)$:

$$\forall k \in \mathbb{N}\,,\ \mathbb{P}\left[Y = k | X = x\right] = \exp(-\mu_Y(x)) \frac{\mu_Y(x)^k}{k!} .$$

Pour ce type de machines, la distribution des âges est donnée par :

x	1	2	3	4
$p_X(x)$	0.1	0.2	0.3	0.4

1. Que représent le paramètre $\mu_Y(x)$? Commenter son expression en fonction de x ?
2. Quelle est la loi du couple (X, Y) ?
3. Je décide d'acheter cette machine d'occasion chez un revendeur qui ne peut m'en spécifier l'âge. Quelle est la distribution du nombre annuel de pannes ?

Exercice 24. Loi conditionnelle. Cas continu.
La hauteur X (en mètres) d'une espèce d'arbres que l'on trouve dans une forêt donnée, dépend de l'âge Y (en années) de l'arbre. Pour un âge donné y, on considère que la hauteur sui une loi normale, dont la moyenne $\mu_X(y)$ et la variance $\sigma_X^2(y)$ dépendent de l'âge. Pour des arbres dont l'âge y varie de 20 à 150 ans, on a :

$$\mu_X(y) = 40\left(1 - \exp(-\frac{y}{50})\right)\ ,\ \sigma_X(y) = \frac{\mu_X(y)}{10}\ .$$

On suppose que la distribution des âges est la loi exponentielle de paramètre 0.02 : $Y \sim \mathscr{E}(0.02)$.

1. Commenter les hypothèses faites (choix des lois, choix des paramètres, ... etc).
2. Quelle est la proportion d'arbres d'âge compris entre 20 et 150 ans ?
3. Quelle est la distribution de la hauteur des arbres âgés de 20 à 150 ans ?
4. Supposons que l'on soit devant un arbre dont la hauteur mesurée vaut x mètres, et dont on sait que l'âge est compris entre 20 et 150 ans. Que peut-on en déduire en ce qui concerne son âge ?
5. Cet exercice est en fait de la statistique. Quel est le paramètre auquel on s'intéresse dans la dernière question ? Montrer que l'on peut mettre cet exercice sous forme d'un modèle bayésien. Quelle est la loi a priori, le modèle statistique et la loi a posteriori dans ce cas ?

Exercice 25. Une urne contient une proportion inconnue p de boules blanches. On y effectue une série de n tirages d'une boule avec remise, et on approche p par la proportion Y_n de boules blanches obtenues sur les n tirages. Montrer que pour tout $\epsilon > 0$, $\mathbb{P}(|Y_n - p| \geq \epsilon) \leq \frac{1}{4n\epsilon^2}$. En déduire une condition sur n pour que l'approximation donne une valeur approchée de p à 0,01 près avec une probabilité supérieure ou égale à 95%.

Exercice 26 (Inégalité de Chernov)**.** Soit ε une variable de signe symétrique ($P(\varepsilon = 1) = P(\varepsilon = -1) = 1/2$) et $(\varepsilon_i)_{i>0}$ une suite de v.a. indépendantes de même loi que ε. On pose $S_n = \sum_{i=1}^n \varepsilon_i$.

1. En appliquant l'inégalité de Markov à $\exp(tS_n)$, montrer que pour tout $t > 0$ et tout $\lambda \geq 0$, $P(S_n > \lambda) \leq \exp(-\lambda t)\text{ch}(t)^n$.
2. En optimisant cette inégalité en t, en déduire que $\forall \lambda \geq 0$,

$$\mathbb{P}\left(\frac{S_n}{n} > \lambda\right) \leq \exp(-nh(\lambda))\,,$$

où la fonction h est définie par

$$h(\lambda) = \begin{cases} \lambda \text{Arcth}(\lambda) + \frac{1}{2}\log(1-\lambda^2) & \text{si } \lambda \in [0;1[\\ +\infty & \text{si } \lambda \geq 1 \end{cases}\ .$$

Exercice 27. Soit X une variable de Cauchy, et $(X_n)_{n\geq 0}$ une suite de v.a.i.i.d de même loi que X. On pose $S_n = \sum_{i=1}^n X_i$. Montrez que S_n/n converge en loi.

Exercice 28. Soit $(X_n)_{n\geq 0}$ une suite de v.a.i.i.d de densité $f(x) = \frac{3}{4}(1-x^2)\ \mathbb{1}_{|x|\leq 1}$. On pose $\xi_n = \max(X_1, \cdots, X_n)$.

1. Montrez que $\xi_n \xrightarrow[n\to+\infty]{proba} 1$.
2. Montrez que $\sqrt{n}(1-\xi_n)$ converge en loi et donner l'expression de la densité de la loi limite.

Exercice 29. On génère à l'ordinateur 100000 nombres aléatoires $u_1, \cdots, , u_{100000}$ selon une loi uniforme sur $[0,1]$, et on calcule leur moyenne géométrique $(u_1 u_2 \cdots u_{100000})^{1/100000}$. Cette valeur sera très proche d'un certain nombre a. Quel est ce nombre a ?

Exercice 30. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires réelles i.i.d. dont la loi commune admet pour densité $f(x) = \mathbf{1}_{[1/2,3/2]}(x)$. On pose pour tout $n \geq 1$, $Y_n = \prod_{i=1}^n X_i$.

1. Montrer que la suite $(Y_n)_{n\geq 0}$ converge presque sûrement vers 0. (*Indication : considérer la suite* $\log Y_n$.)
2. La suite $(Y_n)_{n\geq 0}$ converge-t'elle dans L^1.

Exercice 31. Soient $X_1, \cdots, X_{1000}$ des variables aléatoires suivant la loi uniforme sur l'intervalle $[0,1]$. Soit M le nombre d'entre elles comprises entre 1/4 et 3/4. Déterminer par approximation normale $\mathbb{P}(|M - 500| > 20)$.

Exercice 32. Supposons qu'on ait lancé 10000 fois une pièce de monnaie bien équilibrée.

1. Trouver un intervalle symétrique autour de 5000 contenant le nombre de pile avec une probabilité supérieure à $0,99$.
2. Comparer le résultat avec celui obtenu en appliquant l'inégalité de Chebychev.

Exercice 33. Soit $X \sim \mathscr{P}(\lambda)$ et $Y_\lambda = \frac{X-\lambda}{\sqrt{\lambda}}$.

1. Montrer que Y_λ converge en loi lorsque $\lambda \to \infty$, et déterminer sa loi limite.
2. Retrouver le résultat comme conséquence du théorème Central-Limite , en supposant $\lambda \to \infty$ dans $\mathbb{N}$;
3. Montrer que $e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!} \xrightarrow[n\to\infty]{} 1/2$.